Principal Component Analysis of Thermographic Data

William P. Winfree*[a], K. Elliott Cramer[b], Joseph N. Zalameda[b], Patricia A. Howell[b] and Eric R. Burke[b]

[a]NASA Langley Research Center, MS 225, Hampton, VA 23681
[b]NASA Langley Research Center, MS 231, Hampton, VA 23681

ABSTRACT

Principal Component Analysis (PCA) has been shown effective for reducing thermographic NDE data. While a reliable technique for enhancing the visibility of defects in thermal data, PCA can be computationally intense and time consuming when applied to the large data sets typical in thermography. Additionally, PCA can experience problems when very large defects are present (defects that dominate the field-of-view), since the calculation of the eigenvectors is now governed by the presence of the defect, not the "good" material. To increase the processing speed and to minimize the negative effects of large defects, an alternative method of PCA is being pursued where a fixed set of eigenvectors, generated from an analytic model of the thermal response of the material under examination, is used to process the thermal data from composite materials. This method has been applied for characterization of flaws.

Keywords: Thermography, Nondestructive Evaluation, Principal Component Analysis, Composite

1. INTRODUCTION

Principal component analysis (PCA) has been used extensively as a data reduction technique. Originated by Peason[1] in 1901 and later developed more fully by Hotelling[2] it has been used extensively to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining, as much as possible, the variation present in the data set. For example PCA is used routinely in constructing socio-economic status indices[3], for analysis of spectral imaging data in astrophysics[4] and for facial recognition[5]. Additionally, application of PCA to thermal NDE has also been extensively studied. Typical application of PCA to the reduction of transient thermographic data consists of calculating the principal components of the temporal data through singular value decomposition (SVD) of the experimental data itself. For example, Rajic[6,7] and Valluzzi[8] both use PCA as a contrast enhancement technique for defect detection. Genest[9] and Vavilov[10] provide comparisons between PCA and various other data reduction techniques for defect sizing. Zalameda[11] discusses PCA's use for temporal compression of the thermal data. PCA was used to analyze thermal "flying-spot" data by Hermosilla-Lara[12] for detection of open cracks in metallic specimens. Finally, Marinetti[13] suggests the use of an experimentally derived training set to calculate the principal components. This paper reports on the use of a semi-analytical, one-dimensional model to develop the principal components. The application of this approach will demonstrate how these models are used to quickly reduce and quantify defects in large volumes of thermographic data. Results from the application of this technique to flash IR data, totaling more than 200GB, acquired during a large composite test article inspection are presented. A comparison of this approach with traditional data analysis techniques and a description of the automated large area inspection system are also given.

2. PCA FOR THERMOGRAPHY

In thermographic applications, PCA, which is also referred to as the discrete Karhunen–Loève transform, is a procedure for representing the thermal response as an orthogonal transformation. The orthogonal basis vectors can be based on either a set of measured responses, a set of simulation results or a combination of both. The number of vectors in the orthogonal transform significantly less than the number of points in the discretized thermal response, enabling a representation of the data with significantly reduced number of variables representing the data.

The orthogonal representation of the thermal response is given by

*william.p.winfree@nasa.gov; phone 1 757 864-4963; fax 1 757 -864-8320;

$$\text{[illegible]} \quad (1)$$

where $T_{i,j}(t_n)$ is the digitized thermal response at a given point at of a specimen, $c_{i,j,k}$ are coefficients that represent the thermal response at a given location, $\mu_{i,j}$ is the mean of the response at a given point and $v_k(t_n)$, k=1,…,m are a set of orthogonal vectors where m is less than or equal to the number the number of time steps. The orthogonal vectors are the eigenvectors of the covariance matrix of the thermal responses. The eigenvectors are ordered in such a way that the lowest vectors account for most of the variability in the data. For thermal data typically only a few eigenvectors are required to adequately represent the data. As a result this representation is a good method for both compressing and smoothing the data. The thermal response used to calculate the covariance matrix can be either measured, simulated or combination of both thermal response. A simple way to calculate this covariance matrix is to represent the thermal response at a given point as a column vector T. The covariance matrix (C) is calculated by subtracting the mean for the thermal response and summing the difference times it's transpose for all of the thermal responses or

$$\text{[illegible]} \quad (2).$$

A particular advantage of calculating the covariance matrix this way is a large number of responses can be used to calculate the covariance matrix without performing calculations on an extremely large matrix. A second advantage is the matrix can be stored and updated as more data becomes available.

Often, the dataset being analyzed is used to calculate the covariance matrix. This has the advantage of ensuring the eigenvectors are representative of the data being analyzed. However it also has the disadvantage of the eigenvectors being different for each dataset and the coefficients being difficult to interpret. An alternate method discussed in this paper is to calculate the covariance matrix from a set of model responses. This enables using a consistent set of eigenvectors to represent multiple data sets. A second advantage is the coefficients of the representation can be related to coefficients of the model to provide a more quantitative analysis of the data.

3. DETERMINING EIGENVECTORS FORM TWO-LAYERED SYSTEM TO IMPULSE HEATING

To calculate the covariance matrix, the thermal responses were calculated from a one dimensional model of a two layer system, with a contact resistance between the two layers. This has the advantage of providing thermal responses for a single layer (the contact resistance is very high) or a single layer with a weak contact to another thermal mass, which is the case of a delamination or where there is contact with another large thermal mass.

A simple analytic solution does not exist for the one-dimensional heat flow in a multilayered material. A solution does however exist in Laplace space for two layers of thickness l_1 and l_2 coupled by an intermediate contact resistance(R). Since the configuration of interest is a composite with a delamination, the first and second layers are assumed to have the same thermal conductivity (K) and diffusivity (κ). For the surface with the incident heating, the Laplacian of the front surface temperature response is given by

$$\text{[illegible]} \quad (3)$$

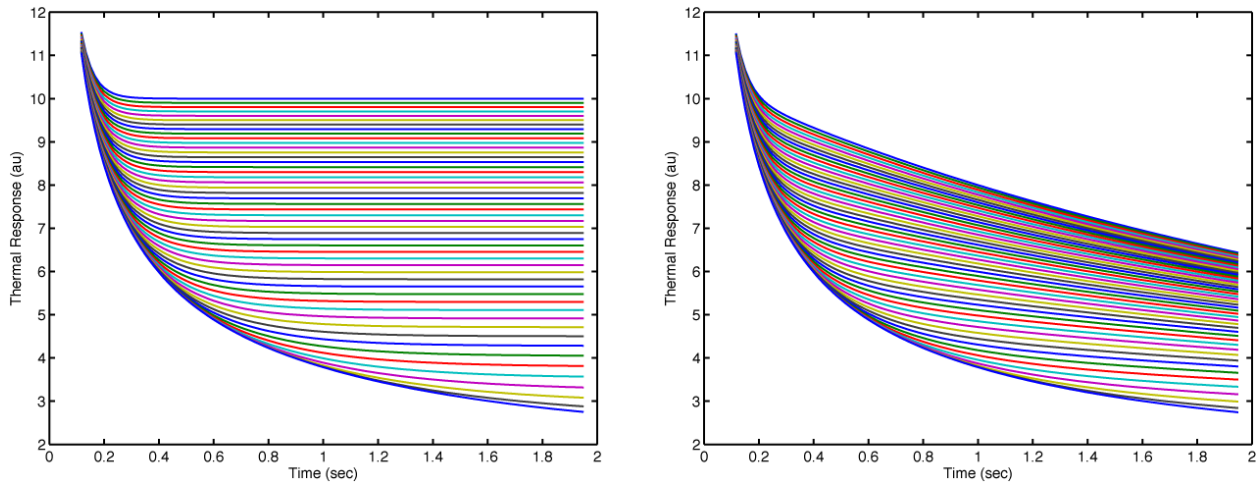
where p is the Laplacian coordinate, q is [illegible] and f is the energy per unit area of the flash. This reduces to the one-dimensional solution of a single layer with a thickness of $l_1 + l_2$ if R equals zero (no contact resistance at the boundary). An advantage of using this to develop a set of eigenvectors for representing thermal responses of three dimensional materials is it can be quickly inverted numerically and for flaws such as delamination and large voids, the thermal response has a shape that can be approximated be a layer that is thermally in contact with another thermal mass.

Thermal responses are calculated to span expected thermal response of the structure. For calculation of the covariance matrix, it has been found to be advantageous to generate a set of thermal responses that do not over represent any particular expected thermal response. For a large contact resistance, this can be done by calculating the thermal response from the

smallest realistic value of κ/l_1^2 (limited by the length of the time record) to the largest realistic value of κ/l_1^2 (limited the first time that the thermal response can be measured) in equal steps. A representative set of equally space responses for a single layer is shown in Fig. 1a.

The contact resistance is reduced until there is significant heat flow from the first layer to the second layer. At this time, no particular method has been established for determining appropriate values for the contact resistance, rather contact resistance are reduced until there is a significant reduction in the long time thermal response of a thin layer to create one set of responses with the first layer thickness being the same as used to create the "single" layer response and then reducing it by a factor of ten and calculating another set. Three different contact resistances (include the extremely large contact resistance which produced a single layer response) were found give a set of eigenvectors that accurately represented all of the acquired data. The set of responses for the contact resistance where there were significant reduction of the long term response of a thin first layer are shown in Fig. 1b. It is as can be seen from a comparison of 1a and 1b, for thicker first layers, the decrease in contact resistance does not have as significant an impact on the response.

The eigenvectors and eigenvalues are calculated numerically from the covariance matrix, using one of the many algorithms available in the public domain or from commercial software packages. The set of eigenvectors and eigenvalues used in this study are shown in Fig. 2. Only the first eigenvector has a shape that qualitatively has the appearance of a thermal response, so clearly there is not straight forward correlation between these shapes and the thickness of the first layer. However, these four eigenvectors are adequate to represent all of the 800 responses generated for the creation of the covariance matrix to better than 1%.



(a) (b)

Figure 1. (a) Simulated thermal responses of single layer used to calculate eigenvectors of representation of experimental thermal responses. (b) Thermal responses of two layer system with contact resistance at the interface between the two layers.

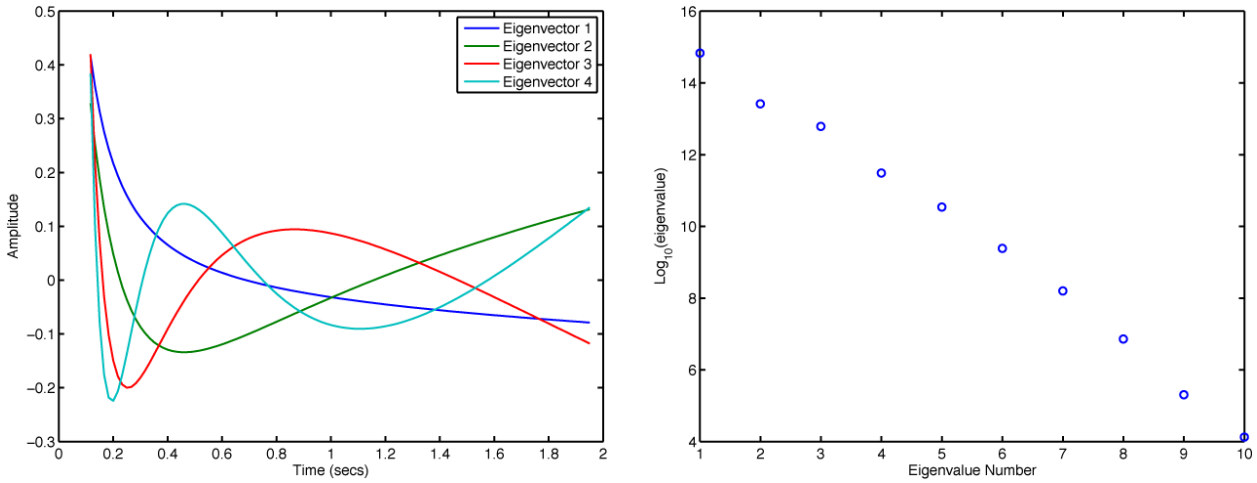
Before calculating the coefficients to represent the thermal response $\boldsymbol{T}_{i,j}$, the mean of the response ($\mu_{i,j}$) is subtracted from the response and the result is normalized by dividing by the square root of the mean of the difference squared or

$$\text{[illegible]} \quad (4),$$

where $a_{i,j}$ is given by

$$\text{[illegible]} \quad (5).$$

The normalization is not necessary for the representation, however, it simplifies the interpretation of the coefficients. The coefficients for representing the thermal response at a given location expressed as a vector $\boldsymbol{c}_{i,j}$ calculated by expressing the eigenvectors as the rows of a matrix times $\mathbf{T'}_{i,j}$.



(a) (b)

Figure 2. (a) First four eigenvectors calculated from model of thermal response (b) Eigenvalues for the first ten eigenvectors.

4. ESTIMATION OF FLAW DEPTH FROM COEFFICIENTS OF EIGENVECTOR RESPRESENTATION

The first four eigenvectors calculated from the two layer response model are used to find coefficients that represent the thermal response at each point on the sample. An estimation of the depth is obtained from the first four coefficients of the eigenvector representation of the measured thermal responses. The sum of squared differences between the coefficients of the measured data and each of the calculated two layer model coefficients is calculated. The depth corresponding to the smallest summed squared differences is taken to be the estimate of the depth. This is similar to a k-means cluster, with the difference being the coefficients for the center of the clusters are determined from the model response, rather than a clustering algorithm.

5. DEMONSTRATION OF TECHNIQUE ON FINITE ELEMENT SIMULATION OF THERMAL RESPONSE

To demonstrate the capability of this technique for determination of the depth of the interface between the two layers, a finite element simulation was performed of 3 flat bottom holes of the different depth in 0.2 cm thick 2 by 6 cm block. The material properties of the layer were a thermal conductivity of 0.97 W/m^{o}K, specific heat of 1200 J/kg/^{o}K and a density of 1600 kg/ m^3. Three holes, with depths of 0.15cm, 0.1cm and 0.05cm and diameters of 1cm. The finite element thermal response at 2 seconds with 1% added noise is shown in Fig. 3(a). As can be seen from that figure, the flaw nearest to the surface is clearly visible, however, the shallowest flat bottom hole is extremely difficult to see in the image. This is in part due to the scaling of the results required to keep the 0.15 cm deep hole from saturating the image.

In Fig. 3(b) the thickness map calculated from coefficients of the first four eigenvectors using the method described in the previous section. In this image, the flat bottom hole that is furthest from the front surface is less visible than the flat bottom hole closest to the surface, however significantly more visible then in the thermal response image. Fig. 4 is a plot of the thicknesses obtained across the center of the flat bottom holes. As can be seen from the figure, there is good agreement between the simulation depths of the holes and the depths estimated from the eigenvector analysis.